AMATH 777

Stochastic Techniques for the Study of Epidemiology

Tom Bury

February 30, 2016

Abstract

Deterministic models have successfully described many epidemiological scenarios, however relatively little progress has been made in a stochastic setting, which is an essential feature for modelling smaller populations. This project introduces two traditional models and analyses them as Birth-Death Processes. Using stochastic techniques, interesting properties of the system are investigated, with results often contradictory to the deterministic counterparts.

1 Motivation

Mathematical modelling in epidemiology has largely been dominated by that of deterministic models. Much insight has been gained through a deterministic approach, however the inherent stochasticity of population dynamics cannot be ignored in many cases. Systems with a high degree of stochasticity can deviate dramatically away from the mean, resulting in poor prediction from a deterministic model. Working with the theory of stochastic processes, one is able to capture the exact dynamics of the probability distributions of the variables, allowing us to quantify the reliability of the deterministic predictions.

When modelling smaller populations, the possibility of extinction becomes important. Consider an endemic disease that is stable in its prevalence due to a constant supply of susceptible individuals. At some point in time, by chance, the disease may fail to be passed on before dying out, resulting in a disease-free population. This may only be captured by embedding stochasticity within the model, and using the methods reviewed in this project, one may also calculate exact probabilities for such events to occur.

The ability to perform heavy simulations is a helpful tool in mimicking the supposed behaviour of a real system. However, a single simulation is not necessarily representative of the average behavioural dynamics, and so a large number of replicate simulations are required to establish confidence. Working analytically (where possible) can therefore be of great benefit, especially when one wishes to obtain precise results for relatively small populations.

This project focuses two of the most basic (yet fundamental) models in epidemiology. We first introduce the terminology along with the deterministic results as predicted by the models. The models are then explored from a stochastic stand-point, from which we view the system as a 'Birth-Death Process'. We then go on to investigate some interesting properties such as extinction times and rates of convergence to equilibrium solutions.

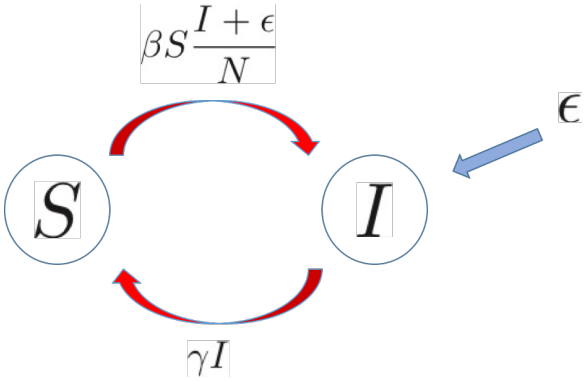


Figure 1: Flow diagram for the SIS Model with imports

2 The SIS Model

2.1 The Basic Framework

In general, modelling in epidemiology is done using compartmental models. This approach assigns a state to each member of a population that corresponds to their state of health. The most simple such model is the SIS model which considers only two states, namely 'Susceptible' and 'Infected'. Individuals in the 'Susceptible' class may become infected if they make direct contact with an individual in the 'Infected' class. Those in the 'Infected' class will eventually recover from the disease at which point they move back into the 'Susceptible' class. Note the SIS model assumes a disease that does not render immunity after recovery from it. Having a 'Recovered' group of immune individuals is a natural extension and is investigated later.

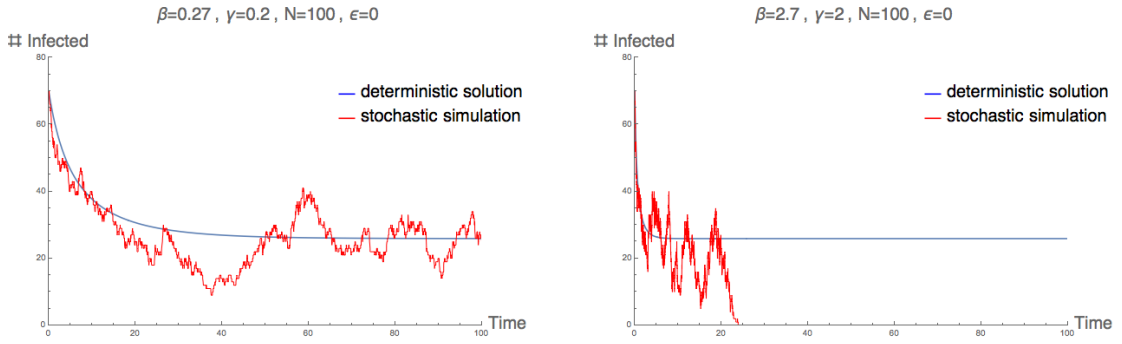
The main features of this model may be captured by a flow diagram (Figure 1) where transition rates have been labelled by the arrows. Parameter values are β, disease transmission rate; γ, recovery rate; N population size; ϵ rate of infectious import from external sources. The deterministic set of equations governing the system are

$$\frac{dS}{dt} = -\beta S\frac{I+\epsilon}{N} + \gamma I \tag{1}$$

$$\frac{dI}{dt} = \beta S\frac{I+\epsilon}{N} - \gamma I. \tag{2}$$

Note that $\frac{d}{dt}(S+I) = 0$ so we are intrinsically assuming that the total population size is constant. This is a reasonable assumption so long as demographic effects are happening on a much slower timescale than that of the disease dynamics, which is often the case. We can then reduce the system to the single first order differential equation

$$\frac{dI}{dt} = \frac{\beta}{N}(N-I)(I+\epsilon) - \gamma I \tag{3}$$



(a) (b)

Figure 2: Stochastic simulations of the SIS model with the same deterministic equilibrium but contrasting transition rates.

(a) For low transition rates, we see fluctuations about the mean with overall shape preserved.
(b) For higher transition rates, the probability of extinction becomes significant, which the deterministic model does not account for.

and find S via $S = N - I$. This equation may be solved analytically and gives rise to a globally stable fixed point $I^* \neq 0$. Thus given any non-zero initial number of infected people $I_0 \in (0, N]$, under these model assumptions I will tend towards the stable equilibrium I^*. Note that if we take $\epsilon \neq 0$ we ensure that $I = 0$ is not an absorbing state.

We should consider the feasibility of this prediction. Population events are random by nature so it is important to identify when the deterministic model is reliable. It turns out that due to the laws of mass action, for very large populations and relatively small transition rates, the deterministic approach yields good predictions. However, we shall consider small population sizes whereby extinction of the disease is a realistic possibility. Simulations for a population size of $N = 100$ demonstrates the significance stochasticity can have (Figure 2). For small transition rates, the overall form is roughly met however the fluctuation sizes are large relative to the mean of the deterministic solution and so shouldn't be ignored. For higher transition rates (common for many diseases), the probability of extinction increases dramatically, rendering the deterministic approximation useless (Figure 2b). A refined approach that takes into account the stochasticity of the problem is clearly required.

2.2 Incorporating Stochasticity into the Model

To incorporate stochasticity into the model, we must observe the underlying processes that are taking place, and their transition rates. In this simple model, there are only two processes, infection and recovery which happen at rates $\frac{\beta}{N}S(I+\epsilon)$ and γI respectively. Let $p_n(t)$ be the probability of n infectious individuals at time t, then modelling this as a 'Birth-Death

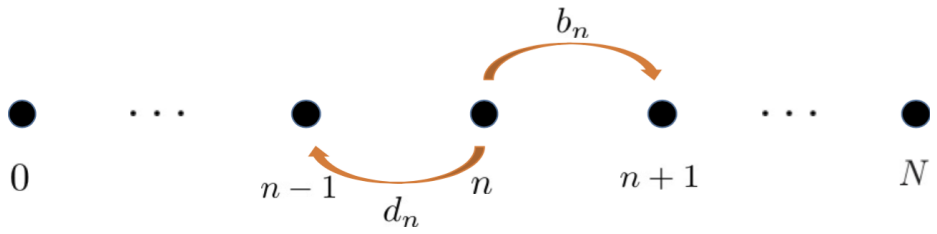


Figure 3: Illustration of a one-dimensional 'Birth-Death' process

Process' (Figure 3), the Kolmogorov forward equations give

$$\frac{dp_n}{dt} = \frac{\beta}{N}(N-n+1)(n-1+\epsilon)p_{n-1} + (\gamma(n+1))\,p_{n+1} - \left(\frac{\beta}{N}(N-n)(n+\epsilon) + \gamma n\right)p_n, \quad (4)$$

where p_{-1} and p_{N+1} are set to zero since the states are not feasible in a population of size N. We may impose the initial condition $p_n(0) = \delta_{n,n_0}$ assuming that we know the initial number of infected individuals in the population to be n_0. The first term of (4) corresponds to moving from the $I = n-1$ state to the $I = n$ state due to an individual becoming infected. Similarly, the second term corresponds to an individual recovering in the $I = n+1$ state, and so the population moves into the $I = n$ state. Finally the last term represents the rate of probability of moving out of the $I = n$ state, either through infection or recovery.

The set of DEs in (4) may be solved numerically as it stands, however further insight may be gained from using matrix formalism. We let $\mathbf{p}$ be the row vector of the probabilities of the system being in each of the $N+1$ states. The Kolmogorov forward equations become

$$\frac{d\mathbf{p}}{dt} = \mathbf{pQ} \quad (5)$$

where $\mathbf{Q}$ is a tridiagonal matrix consisting of the transition rates. A convenient property of the Kolmogorov forward equations is that they give rise to a linear set of DEs, despite the more complex non-linear transition probabilities. The solution to (5) is characterised completely by the eigenvalues and eigenvectors of the matrix $\mathbf{Q}$. The simplest formulation of the solution is as we know

$$\mathbf{p}(t) = \mathbf{p}(0)\exp(\mathbf{Q}t) \quad (6)$$

however, as opposed to exponentiating a matrix, we may decompose the solution as

$$\mathbf{p}(t) = \sum_{n=1}^{N+1} q_n \exp(\lambda_n t)\mathbf{l}_n, \quad (7)$$

where λ_n and $\mathbf{l}_n$ are the eigenvalues and left-eigenvectors respectively of $\mathbf{Q}$. The values for q_n may be found from the initial probability distribution using $q_n = \mathbf{r}_n.\mathbf{p}(0)$ where $\mathbf{r_n}$ are the right-eigenvectors of $\mathbf{Q}$. We also assume that the $\mathbf{l}_n$ are normalised so that they are valid probability distributions and the eigenvalues are ordered according to the size of their real

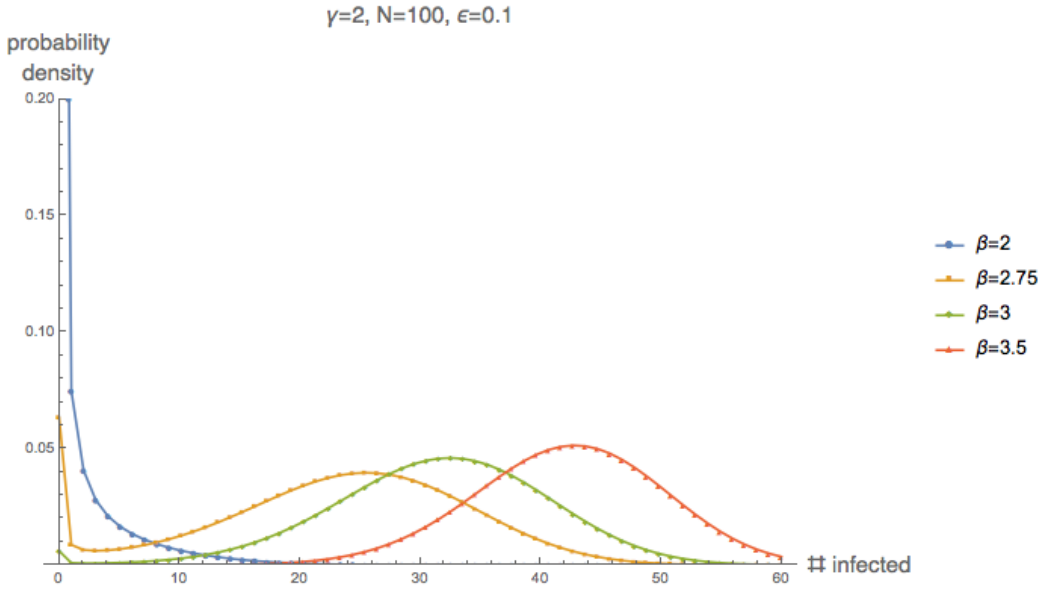


Figure 4: Stationary probability distributions of the SIS model for various β values. (Colour version available online.)

part $(0 = \lambda_1 \geq \Re(\lambda_2) \ldots \geq \Re(\lambda_{N+1}))$. Analysing the matrix $\mathbf{Q}$ for the SIS model, one finds that $\lambda_1 = 0$ and all other eigenvalues are real and negative. Thus the long-term distribution of the stochastic SIS model is simply given by $\mathbf{l}_1$.

I have run this calculation for various parameter values and plotted the distributions in Figure 4. For low disease transmission rate β, there is a high probability of disease extinction as one would expect. For higher β, we notice a Gaussian shape distribution centred around the deterministic equilibrium point. Of particular interest are the intermediate cases around $\beta = 2.75$, giving rise to bi-modal distribution. As well as having a Gaussian shape distribution around the deterministic equilibrium, there is a sharp peak at zero, representing a significant probability of extinction, completely ignored by a deterministic model.

Of course we can gain more than just the long-term behaviour from (7). The other eigenvalues give us information on the rate of convergence to the equilibrium distribution. In particular, for the SIS model, one finds that the order of magnitude jumps significantly from λ_2 to λ_3. Thus contributions from eigenvalues λ_3 to λ_{N+1} decay much more rapidly. What's more is that the entry of the eigenvector $\mathbf{l}_2$ corresponding to the disease-free state $I = 0$ is significantly larger than the other entries. So if the initial disease state is near extinction, the value of q_2 will be close to one, and this contribution will decay relatively slowly. In other words, there is a much slower convergence to equilibrium from a state of extinction, than from any infected state as one would expect.

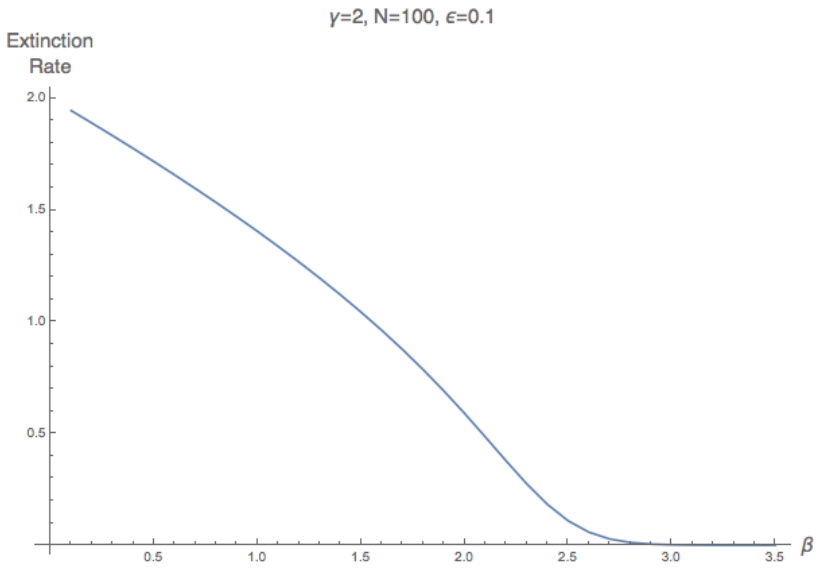


Figure 5: Disease extinction rate for various levels of disease infectivity β

A useful property to calculate is the rate of extinction from the equilibrium state, which is simply given by

$$\rho_{ex} = p_1^* \gamma, \tag{8}$$

i.e. the probability of there begin a single infected individual who then recovers. The mean time to extinction is then simply

$$\tau_{ext} = \frac{1}{\rho_{ext}}. \tag{9}$$

In plotting these rates (Figure 5), we find that for $\beta > 3$ the rate of extinction is close to zero (as confirmed in Figure 4) and as β decreases below this value the extinction rate increases in an approximately linear fashion.

3 The SIR Model

Should we wish to model a disease that individuals become immune to after recovery, the SIR model is the natural first choice. The methods applied previously work equally well however now there are now two independent variables making the problem more complex. Assuming equal birth and death rates the deterministic model is

$$\frac{dS}{dt} = dN - \beta S(I+\epsilon) - dS, \tag{10}$$

$$\frac{dI}{dt} = \beta S(I+\epsilon) - \gamma I - dI, \tag{11}$$

$$\frac{dR}{dt} = \gamma I - dR \tag{12}$$

where the equation for R is essentially redundant since it may be obtained from $S+I+R = N$. There are now several transitions that need to be accounted for (birth, death of a susceptible individual, death of an infected individual, infection and recovery). Incorporating these into the Kolmogorov forward equations gives

$$\begin{aligned}\frac{dp_{S,I}}{dt} &= [\beta(S+1)(I-1+\epsilon)]p_{S+1,I-1} + [\gamma(I+1)]p_{S,I+1} \\ &+ [d(N-(S-1)-I)]p_{S-1,I} + [d(I+1)]p_{S-1,I+1} \\ &- [\beta S(I+\epsilon) + \gamma I + d(N-S-I) + dI]p_{S,I}\end{aligned} \tag{13}$$

where $p_{S,I}$ is the probability of being in the state with S susceptible and I infected individuals. We may again write this in matrix form by mapping the two-dimensional $p_{S,I}$ onto a one-dimensional vector $\mathbf{p}$. One again finds that the first eigenvalue λ_1 is zero, while all the others have negative parts. Thus the long-term distribution is again given by $\mathbf{l}_1$. Figure $6(a)$ graphs the probability distribution along with plotted points taken from a simulation. Since the simulation is taken over a long time period, and many points are taken, we see strong agreement between the analytical results and those from simulation.

The second eigenvalue and eigenvector again describe the slow escape from the disease free state ($S = N, I = 0$) with $\lambda_2 \approx -0.3468$. Interestingly, the deterministic model again fails here by predicting strong instability around the disease free state. One can show from (10,11,12) that for small $I(0)$ and early times

$$I(t) \approx I(0)e^{(R_0-1)\gamma t} \tag{14}$$

where $R_0 = \frac{\beta N}{\gamma+d}$ is the basic reproductive ratio. Inserting parameter values gives a growth rate of $\exp(2.333t)$. The difference may be attributed to the fact that the stochastic model takes into account the possibility of stochastic extinctions and reintroduction via import, adding the the escape time.

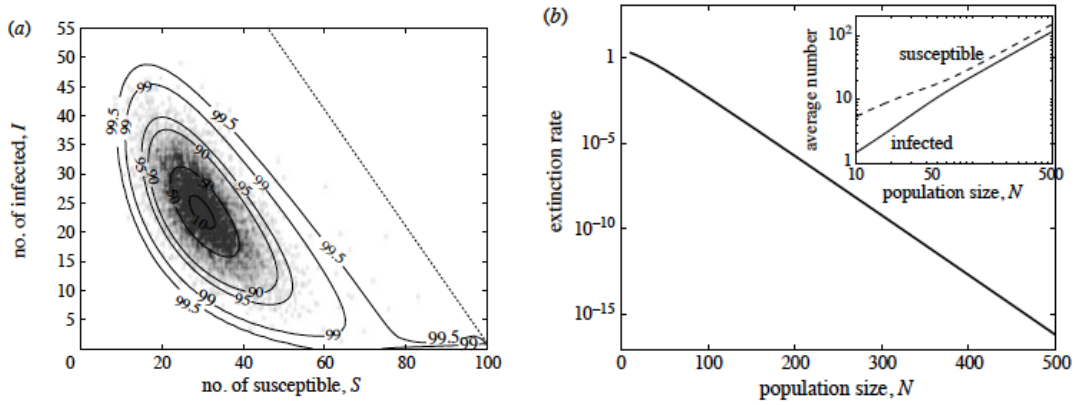


Figure 6: Dynamics of the stochastic SIR model with strong demography [1].
(*a*) Illustration of the level curves from the long-term probability distribution given by $\mathbf{l}_1$. The dots are from a simulation taken over a long period of time. Darker areas represent states that were visited more frequently, showing agreement with simulation and analysis.
(*b*) Extinction rates from the stationary distribution for various population sizes.

The rate of extinction may be calculated in a similar to way to before however now we must consider all possible states with a single infected individual and sum over them. There is also the possibility that this individual dies as opposed to recovering, equally causing extinction of the disease. Thus the rate is given by

$$\rho_{ex} = \sum_{S=0}^{N-1} p^*_{S,1}(\gamma + d) \tag{15}$$

and plotted for various population sizes in Figure 6(b). It is found that the extinction rate decreases exponentially with population size. The inset shows the average number of infected and susceptible individuals which increases roughly linearly with population size, as one would expect.

4 Discussion

This project has demonstrated methods for the exact evaluation of the evolution of a stochastic population model. Working using the Kolmogorov forwards equations on two of the most famous disease models in epidemiology, we gain insight into the stochastic behaviour inherently posed by the random nature of population events. It has been shown using matrix formulation that stationary probability distributions may be calculated with relative ease, along with other interesting properties such as time to extinction and convergence rates.

Despite the power of this approach, it is limited to use on modelling relatively small populations in order that computation time does not become excessive. In any case, it is for small populations that stochasticity plays the most significant role, especially when considering the possibility of extinction. This approach would therefore be well-suited to modelling disease spread within farms or hospital wards, for example.

We finish by outlining the key benefits to this analysis over the use of simulation. Most importantly, we only need perform a single calculation (from the Kolmogorov Forward equations) to find $\mathbf{p}(t)$, which describes the dynamics of an infinite ensemble of stochastic realisations. In particular, when one requires precise probabilities of rare events, the number of simulations required is very large and would be an inefficient approach. One may also be interested in rates of convergence to equilibrium distributions, which may be calculated using theory from dynamical systems on the set of coupled ordinary differential equations generated by the Kolmogorov forward equations. Finally, one may consider a range of initial conditions very efficiently once the initial calculation has been performed. I'm sure these techniques play an important role in future modelling of disease outbreak in small populations.

References

[1] Keeling M.J., Ross J.V. On methods for studying stochastic disease dynamics. *J. R. Soc. Interface* (2008) : 171-181

[2] Keeling M.J. Simple stochastic models and their power-law type behaviour. *Theoretical Population Biology* (2000) 58: 21-31.

[3] Gillespie, D.T. A general method for numerically simulating the stochastic evolution of coupled chemical reactions. *J. Comput. Phys.* (1976) 403-434.

[4] Miskovic Z. L. AMATH 777 Course Notes